MEPS HC-081:
1996-2001 Risk Adjustment Scores
Public Use File

May 2006

TABLE OF CONTENTS

A. Data Use Agreement

Individual identifiers have been removed from the micro-data contained in the files that are part of this Public Use Release. Nevertheless, under sections 308 (d) and 903 (c) of the Public Health Service Act (42 U.S.C. 242m and 42 U.S.C. 299 a-1), data collected by the Agency for Healthcare Research and Quality (AHRQ) and /or the National Center for Health Statistics (NCHS) may not be used for any purpose other than for the purpose for which they were supplied; any effort to determine the identity of any reported cases, is prohibited by law.

Therefore in accordance with the above referenced Federal Statute, it is understood that:

No one is to use the data in this data set in any way except for statistical reporting and analysis; and

If the identity of any person or establishment should be discovered inadvertently, then (a) no use will be made of this knowledge, (b) The Director Office of Management AHRQ will be advised of this incident, (c) the information that would identify any individual or establishment will be safeguarded or destroyed, as requested by AHRQ, and (d) no one else will be informed of the discovered identity.

No one will attempt to link this data set with individually identifiable records from any data sets other than the Medical Expenditure Panel Survey or the National Health Interview Survey.

By using this data you signify your agreement to comply with the above stated statutorily based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates Title 18 Part 1 Chapter 47 Section 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

The Agency for Healthcare Research and Quality requests that users cite AHRQ and the Medical Expenditure Panel Survey as the data source in any publications or research based upon these data.

B. Background

This documentation describes one in a series of public use files from the Medical Expenditure Panel Survey (MEPS). The survey provides a new and extensive data set on the use of health services and health care in the United States.

MEPS is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian non-institutionalized population. MEPS is cosponsored by the Agency for Healthcare Research and Quality (AHRQ) and the National Center for Health Statistics (NCHS).

MEPS comprises three component surveys: the Household Component (HC), the Medical Provider Component (MPC), and the Insurance Component (IC). The HC is the core survey, and it forms the basis for the MPC sample and part of the IC sample. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

MEPS is the third in a series of national probability surveys conducted by AHRQ on the financing and use of medical care in the United States. The National Medical Care Expenditure Survey (NMCES, also known as NMES-1) was conducted in 1977, the National Medical Expenditure Survey (NMES-2) in 1987. Beginning in 1996, MEPS continues this series with design enhancements and efficiencies that provide a more current data resource to capture the changing dynamics of the health care delivery and insurance system.

The design efficiencies incorporated into MEPS are in accordance with the Department of Health and Human Services (DHHS) Survey Integration Plan of June 1995, which focused on consolidating DHHS surveys, achieving cost efficiencies, reducing respondent burden, and enhancing analytical capacities. To accommodate these goals, new MEPS design features include linkage with the National Health Interview Survey (NHIS), from which the sampled households for the MEPS HC are drawn, and continuous longitudinal data collection for core survey components. The MEPS HC augments NHIS by selecting a sample of NHIS respondents, collecting additional data on their health care expenditures, and linking these data with additional information collected from the respondents' medical providers, employers, and insurance providers.

1.0   Household Component

The MEPS HC, a nationally representative survey of the U.S. civilian non-institutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment.

The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a 2½ - year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection

rounds is launched each year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures.

The sample of households selected for the MEPS HC is drawn from among respondents to the NHIS, conducted by NCHS. The NHIS provides a nationally representative sample of the U.S. civilian non-institutionalized population, with oversampling of Hispanics and blacks.

2.0 Medical Provider Component

The MEPS MPC supplements and/or replaces information on medical care events reported in the MEPS HC by contacting medical providers and pharmacies identified by household respondents. The MPC sample includes all home health agencies and pharmacies reported by HC respondents. Office-based physicians, hospitals, and hospital physicians are also included in the MPC but may be subsampled at various rates, depending on burden and resources, in certain years.

Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents. The MPC is conducted through telephone interviews and record abstraction.

3.0 Insurance Component

The MEPS IC collects data on health insurance plans obtained through employers, unions, and other sources of private health insurance. Data obtained in the IC include the number and types of private insurance plans offered, benefits associated with these plans, premiums, contributions by employers and employees, eligibility requirements, and employer characteristics.

Establishments participating in the MEPS IC are selected through four sampling frames:

- A list of employers or other insurance providers identified by MEPS HC respondents who report having private health insurance at the Round 1 interview.

- A Bureau of the Census list frame of private sector business establishments.

- The Census of Governments from Bureau of the Census.

- An Internal Revenue Service list of the self-employed.

To provide an integrated picture of health insurance, data collected from the first sampling frame (employers and insurance providers) are linked back to data provided by the MEPS HC respondents. Data from the other three sampling frames are collected to provide annual national and State estimates of the supply of private health insurance available to American workers and to evaluate policy issues pertaining to health insurance.

The MEPS IC is an annual panel survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone followup for nonrespondents.

4.0 Survey Management

MEPS data are collected under the authority of the Public Health Service Act. They are edited and published in accordance with the confidentiality provisions of this act and the Privacy Act. NCHS provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports and microdata files. Summary reports are released as printed documents and/or electronic files on the MEPS web site (www.meps.ahrq.gov). All microdata files are available for download from the MEPS web site in compressed formats (zip and self-extracting executable files.) Selected data files are available on CD-ROM from the MEPS Clearinghouse.

For printed documents and CD-ROMs that are available through the AHRQ Publications Clearinghouse, write or call:

AHRQ Publications Clearinghouse
Attn: (publication number)
P.O. Box 8547
Silver Spring, MD 20907
800/358-9295
410/381-3150 (callers outside the United States only)
888/586-6340 (toll-free TDD service; hearing impaired only)

Be sure to specify the AHRQ number of the document or CD-ROM you are requesting.

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Financing, Access and Cost Trends, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850 (301/427-1406).

C. Technical Information

1.0   Data File Contents

This documentation describes the 1996-2001 Relative Risk Scores Public Use File derived from the respondents to the Medical Expenditures Panel Survey (MEPS) sample for Panels 1 through 5. To obtain analytic variables, the records on this file must be linked to the corresponding MEPS public use data sets by the sample person identifier (DUPERSID).

This file contains a total of 75,275 persons. Each record contains a PANEL indicator which identifies the time period the respondent was in the survey. For example if PANEL=1, the respondent was in the MEPS survey for 1996 and 1997.

2.0   Relative Risk Scores based on the DCG Model in MEPS

A large literature describes methods for estimating the relative propensity to consume health services. These methods are used to adjust for the risk of future utilization when predicting or explaining health care utilization and costs. These "risk adjustment" methods are typically based on diagnostic information from claims data. One well known risk adjustment model, the DCG model, has been developed by researchers at DxCG Inc.

The DCG model takes diagnostic information, based on claims data, and aggregates specific diagnoses into broader clinically meaningful categories. DxCG's Hierarchical Condition Categories (HCCs) are based on the 5-digit level ICD-9-CM diagnosis codes. Each code is classified into one of 184 condition categories, and hierarchies are further imposed to make predictions more robust to variations in how disease codes are captured, to reward specific coding, and to increase model stability. Regression models have been developed using large national samples to predict various outcomes. Age, sex, HCCs and interaction terms are included in the models. The individual-level prediction is a relative risk score (RRS). The relative risk score is a summary of disease burden and expected annual health care resource use at the individual level. The RRS can be converted into a dollar prediction by multiplying by an appropriate sample mean. For example, if a reference population has $2000 mean costs, then multiply RRS by $2000. HCC/DCG models are described in several articles referenced at the end of this note.

DCG prospective RRSs are also good "generic" measures of disease burden. Studies have shown that people with higher RRS scores go on to use more hospitalizations, ER services and home care, and to experience higher mortality. These scores are widely employed in health policy studies, budgeting, payment, pricing, negotiation, provider profiling, disease management reconciliation, and resource planning.

The MEPS public use data contain 3-digit level ICD-9-CM diagnosis codes. To add value

for health services researchers, AHRQ asked DxCG Inc. to use these diagnosis codes to generate a relative risk score for each individual respondent, to enable risk-adjustment for examining future health care spending, and as a general proxy for morbidity due to disease burden.

First, DxCG examined how using 3-digit diagnoses (rather than 5-digit codes) would affect the prospective DCG/HCC model's performance. They concluded that, although using 3-digit codes would reduce the model's specificity in clinical classification and its predictive accuracy, the loss in specificity and predictive power was small. Therefore, a DCG/HCC prospective model was deemed to be applicable to the MEPS data using 3-digit ICD-9 codes.

Insurance coverage presented a second complication in applying DCG models to the MEPS data. DCG models have been developed using linear regression on large national claims datasets from particular insurers. Different models have been developed for different datasets: One risk adjustment model was derived for Medicare claims, another for claims for privately insured individuals, and a third for Medicaid claims data. While the majority of MEPS respondents have one source of insurance coverage during a calendar year, people can be uninsured, and they can change insurance coverage during a year. To accommodate this complexity, we developed a variable that represents the predominant form of coverage for each respondent. This variable, INSCAT1, has four categories:

1	Medicare
2	Private
3	Medicaid
4	Uninsured

Respondents were assigned to a category based on the number of months of each type of coverage (or no coverage). Thus, if someone had 7 months of private coverage and 5 months of Medicare, the person was coded as private (INSCAT1 = 2). If someone had equal months of coverage for two or more different sources, their classification was based on the following hierarchy: Medicare, private, Medicaid, uninsured.

> The DCG models were developed to predict health care costs. **Note that costs refer to the kinds of costs covered within an insurance system.** Thus, for example, a person with high long term care costs may look less expensive to a Medicare model (since Medicare does not pay long term care costs) than he or she would to a Medicaid model (which does pay such costs).

To provide maximum flexibility and information for users of MEPS data, each of the three established DCG prediction models (Medicare, private, and Medicaid) was applied to each MEPS respondent, regardless of the person's insurance status. Within each type of DCG model (Medicare, private, and Medicaid) there are two model specifications: A basic model included only information on the person's age and sex ("age/sex" or "A/S"

model), and a more elaborate model also included information on the HCCs, based on medical conditions reported for each respondent in MEPS. Thus, **six** relative risk scores, based on a combination of model type (Medicare, private, and Medicaid) and model specification ("A/S" only or age/sex and HCCs), have been produced for each person.

Note: Risk scores were not derived for respondents who were not eligible for the first year of a panel (e.g., persons born in Year 2, persons moving into a household in Year 2).

Table 1 shows the variable names, corresponding to the models used to implement the DCG prediction, and the inputs used in each model.

Table 1 – Prospective DCG Relative Risk Scores in MEPS

DCG Risk Score Name (in DxCG, Inc. software)	Model Type*	YEARONE Model Inputs
RRSASMC	A/S_Medicare	Age, Sex
RRSHCCMC	HCC_Medicare	Age, Sex, Diagnoses
RRSASPV	A/S_Private	Age, Sex
RRSHCCPV	HCC_Private	Age, Sex, Diagnoses
RRSASMD	A/S_Medicaid	Age, Sex
RRSHCCMD	HCC_Medicaid	Age, Sex, Diagnoses, Eligibility Categories

* "A/S" refers to models based on age and sex alone. "HCC" stands for the Hierarchical Condition Category modeling framework that organizes diagnostic information into profiles, which, in conjunction with demographic data, are used (in these prospective models) to predict next year's health care cost. The second part of each type name refers to the population on which the model was originally derived: Medicare, commercially (privately) insured, or Medicaid.

DCG models were applied to MEPS respondents in Panels 1-5. Each panel covers a two-year observation period. In all cases, a prospective DCG model was used, which estimated relative risks in a subsequent year, based on the current year's information. In MEPS, each panel is observed for a two-year period. Thus, the DCG models were based on medical condition information in the first year of a panel to predict relative expenditures in the second year. In the models, age was coded as age in the second year of a panel.

Normalization

Risk scores are "made relative" by multiplying by a normalizing constant, chosen so that the scores average to 1.00 within specified MEPS subpopulations. Thus, relative risk scores (RRSs) are normalized, positive predictions of future (prospective) total health care spending, where a score of 1 refers to a person whose expected costs next year are "average" in a specified population. Regardless of how they are normalized, relative risk scores convey relative expected costliness, so that, when applying the same model to any group of people under a given type of health care benefit, RRS = 1.5 indicates expected costs 50% higher than RRS = 1.0.

For the MEPS data, a separate normalization was performed for each combination of panel and INSCAT1. Table 2 is based on the standard RRSs produced by the DCG modeling software, for each combination of panel and INSCAT1, prior to normalization. The entries are the mean RRS for each cell; in calculating the mean, data were weighted by the analytic weight derived for longitudinal analyses of each panel (LONGWGT). For example, we applied DxCG's HCC Medicare model to all members of MEPS panels 1 through 5, producing the standard Medicare relative risk scores. The mean of these scores, among only the (n = 2,566) people in the MEPS panel 1 subgroup with INSCAT1 = Medicare, was calculated as 0.5605350. (See Table 2.) Similarly, the mean standard Medicare relative risk score among only the (n = 2,566) people in the MEPS panel 2 subgroup with INSCAT1 = Medicare was 0.5726652.

Table 2 -- Conversion Factors (Numbers needed to multiply by to recover the original DxCG-model risk score values)

DCG/HCC Model

Panel	Private	Medicare	Medicaid
1	0.9383405	0.5605350	0.3576319
2	0.9147182	0.5726652	0.3805723
3	0.8844287	0.5538300	0.3602353
4	0.8687499	0.5554040	0.3573271
5	0.8973654	0.5351716	0.3298320

Age/Sex Model

Panel	Private	Medicare	Medicaid
1	0.9235053	1.0033810	0.5848393
2	0.9330559	0.9991936	0.5849619
3	0.9373458	1.0008386	0.5546255
4	0.9455516	1.0014055	0.5543329
5	0.9497323	1.0019651	0.5374621

The mean standard RRSs were then used to normalize the individual relative risk scores, by panel and INSCAT1. Thus, all panel 1 relative risk scores based on the DCG Medicare model (n = 19,529, including everyone in panel 1, regardless of insurance) were divided by 0.5605350 to produce the variable labeled RRSHCCMC for panel 1. Similarly, all Medicare relative risk scores in panel 2 were divided by 0.5726652, to create the RRSHCCMC score for panel 2. Thus, the average RRSHCCMC score for panel 1 people in Medicare (INSCAT1=1) is 1, and the average RRSHCCMC score for panel 2 people in Medicare is also 1. This process was repeated for each of the other panels. The overall process was then repeated for the DCG private model, yielding the variable RRSHCCPV, and for the Medicaid model, yielding RRSHCCMD.

In other words, within each combination of panel and INSCAT1, the average risk score is normalized to 1.000. This allows researchers to conduct analysis by panel or by insurance coverage type across panels or both. (Thus, while the average RRSHCCMC score for people in Medicare in panel 1 will equal 1.0, the average RRSHCCMC score for persons

with private insurance or with Medicaid in panel 1 will not equal 1. However, the mean RRSHCCPV score will equal 1.0 for persons with private coverage in panel 1, but will not equal 1 for persons with Medicare or Medicaid in panel 1.)

If a researcher wants to convert the relative risk scores to dollar predictions, he/she needs to multiply the average expenditure for a combination of panel and INSCAT1 by the relative risk score for that combination. To move from a relative prediction to a dollar prediction for a person in any of these three insured populations, multiply the risk scores by the average expenditure for the corresponding panel*INSCAT1, as given in Table 3.

The HCC private insurance model predicts subsequent costs best (in terms of R-squared) for the uninsured. To create dollar predictions (that match the observed costs) for an uninsured respondent in a panel, you can multiply the RRSHCCPV relative risk score for an uninsured respondent in a panel by the mean observed cost for uninsured respondents in that panel.

Table 3 – Average Expenditure by Panel and Insurance Category (INSCAT1)

Panel	Private	Medicare	Medicaid	Uninsured
1	1,546.96	5,310.10	1,180.66	$980.05
2	1,387.46	6,030.48	1,302.50	$869.94
3	1,681.35	5,352.68	1,550.88	$1,213.00
4	1,775.26	6,303.31	1,789.67	$1,116.85
5	1,869.04	6,980.15	1,366.49	$1,355.17

DxCG BIBLIOGRAPHY
Publications by DxCG Senior Scientists

Zhao Y, A Ash, RP Ellis et al	"Predicting Pharmacy Costs and Other Medical Costs Using Diagnosis and Drug Claims." <u>Medical Care</u> 43 (1): 34-43, January 2005.
Pope G, J Kautter, RP Ellis, A Ash, J Ayanian, et al	"Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC model." <u>Health Care Financing Review</u>. Summer 2004.
Ellis RP, MS Kramer, JF Romano, R Yi	"Applying Diagnosis-based Predictive Models to Group Underwriting." <u>Health Section News</u>; August 2003. 1, 4-8.
Zhao Y, A Ash, J Haughton, B McMillan	"Identifying Future High-Cost Cases Through Predictive Modeling." <u>Disease Management and Health Outcomes</u> 2003; 11(6): 389-397.
Zhao Y, A Ash, RP Ellis, et al	"Disease burden profiles An Emerging Tool for Managing Managed Care." <u>Health Care Management Science</u> (2002); 5(3) 211-219.
Shen, Y, RP Ellis	"Cost-Minimizing Risk Adjustment." <u>Journal of Health Economics</u>. (2002) 21(3) pp 515-530.
Shen, Y, RP Ellis	"How Profitable is Risk Selection? A Comparison of Four Risk Adjustment Models." <u>Health Economics</u>. (2002) 11(2) 165-174.
Ash, A, Y Zhao, RP Ellis, MS Kramer	"Finding Future High-cost Cases Comparing Prior Cost Versus Diagnosis-based Methods." <u>Health Services Research</u> 36(6) Part II December 194-206. (2001).
Zhao Y, Ellis RP, A Ash, et al	"Measuring Population Health Risks Using Inpatient Diagnosis and Outpatient Pharmacy Data". <u>Health Services Research</u>. 36(6) Part II December 180-193. (2001).
Ash, A, F Porell, L Gruenberg, et al	"Adjusting Medicare Capitation Payments Using Prior Hospitalization." <u>Health Care Financing Review</u> 10(4) 17-29, 1989.
Ellis, RP, A Ash	Refinements to the Diagnostic Cost Group Model. <u>Inquiry</u> 32 1-12, Winter 1995.
Ellis, RP, G Pope, et al	"Diagnosis-Based Risk Adjustment for Medicare Capitation Payments." <u>Health Care Financing Review</u>, Spring 1996.
Pope, G, CF Liu, RP Ellis, A Ash et al	"Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment." <u>Health Care Financing Review</u>, (2000) Spring 21 (3) 93-118.
Pope, G et al	"Evaluating Alternative Risk Adjusters for Medicare," <u>Health Care Financing Review</u>, 1998.